

# CONSTRUÇÃO DE MODELOS *CREDIT SCORING* COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

**Elaine Aparecida Araújo**

[elaineadc@yahoo.com.br](mailto:elaineadc@yahoo.com.br)

UNIVERSIDADE FEDERAL DE PERNAMBUCO - UFPE / Brasil

**Charles Ulises de Montreuil Carmona**

[ccarmona@uiuc.edu](mailto:ccarmona@uiuc.edu)

UNIVERSIDADE FEDERAL DE PERNAMBUCO – UFPE / Brasil

## RESUMO

Os modelos de *Credit Scoring* são modelos quantitativos empregados comumente por instituições financeiras na mensuração e previsão do risco de crédito, possuindo uso consolidado no processo de concessão de crédito destas instituições. O presente trabalho objetivou avaliar a possibilidade de aplicação de modelos *Credit Scoring* em uma instituição de microcrédito denominada Fundo Rotativo de Ação da Cidadania – Cred Cidadania. Para isso, foram coletados dados relativos a uma amostra de clientes do Cred Cidadania, e estes dados foram utilizados para desenvolver dois tipos de modelos de *Credit Scoring*: um de aprovação de crédito e um outro chamado *behavioural scoring* (escoragem comportamental). As técnicas estatísticas empregadas na construção dos modelos foram análise discriminante e regressão logística. Os resultados do estudo demonstraram que os modelos *Credit Scoring* obtêm desempenho satisfatório quando utilizados na análise de risco de crédito na instituição de microcrédito Cred Cidadania, alcançando um percentual de classificação correta dos clientes de cerca de 80%. Os resultados indicam também que o uso de modelos *Credit Scoring* fornece subsídios à instituição, auxiliando-a na prevenção e redução da inadimplência e na diminuição dos seus custos operacionais, dois problemas que afetam a sua sustentabilidade financeira.

**Palavras-Chave:** Risco de Crédito. Modelos Credit Scoring. Microcrédito. Análise Discriminante. Regressão Logística.

## THE DEVELOPMENT OF CREDIT SCORING MODELS WITH LOGISTIC REGRESSION AND DISCRIMINANT ANALYSIS FOR THE CREDIT RISK MANAGEMENT OF A MICROCREDIT INSTITUTION

### ABSTRACT

The Credit Scoring models are quantitative ones usually used by financial institutions in measure and credit risk forecast, owning consolidated application during the credit concession process of these institutions. This work objectives to evaluate the possibility of Credit Scoring models application in a microcredit institution denominated Fundo Rotativo de Ação  
REAd – Edição 62 Vol 15 N° 1 jan-abril 2009

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

da Cidadania – Cred Cidadania. For this purpose, some data relative to a clients' sample Cred Cidadania were collected and used to develop two Credit Scoring model types: one relating to credit approval and another one named behavioural scoring. The statistical technique used in the models construction was logistic regression. The study results demonstrated that Credit Scoring models obtain satisfactory performance when used in the Cred Cidadania microcredit institution credit risk analysis, as supporting instruments to rely this process. The results also indicate that Credit Scoring models application supplies subsidies to the institution, assisting it in the prevention and reduction of its insolvency as in the decrease of its operational costs, two problems that affect its financial sustainability.

**Key Words:** Credit Risk. Credit Scoring Models. Microcredit. Discriminant Analysis. Logistic Regression.

## 1. Introdução

Mudanças ocorridas no cenário financeiro mundial, a partir dos anos 90, tais como, desregulamentação das taxas de juros e câmbio, aumento de liquidez e aumento da competição bancária, fizeram com que as instituições financeiras se preocupassem cada vez mais com o risco de crédito.

A necessidade de controle e gerenciamento eficaz do risco fez com que as instituições financeiras passassem a primar pelo aperfeiçoamento das técnicas utilizadas para essa função, com o desenvolvimento de inúmeros modelos quantitativos pelas próprias instituições financeiras e pelas empresas de consultoria.

Tendo em vista que os modelos quantitativos já possuem uso consolidado na análise de crédito de instituições financeiras tradicionais, avaliou-se, neste trabalho, a possibilidade de aplicação dos mesmos na metodologia de avaliação de risco de crédito nas instituições de microcrédito.

Na definição de Barone et al (2002), o microcrédito é a concessão de crédito de baixo valor a pequenos empreendedores informais e microempresas sem acesso ao sistema bancário tradicional.

A metodologia adotada nos programas de microcrédito é direcionada para o perfil e necessidade dos tomadores, e tem como característica o acompanhamento ao crédito concedido, através da presença de profissionais especializados que avaliam a evolução do empreendimento bem como da sua capacidade de pagamento, além de prestarem assistência na gestão do negócio.

As principais vantagens da metodologia adotada no microcrédito referem-se ao baixo custo que proporcionam aos clientes, tendo como características principais a proximidade do cliente, mínimo de burocracia, agilidade na entrega do crédito.

Por outro lado, as instituições de microcrédito possuem altos custos operacionais, baixas escalas e altas taxas de inadimplência, decorrentes, principalmente, da má qualidade de suas carteiras.

A conjugação desses e outros problemas afetam a sustentabilidade financeira das instituições de microcrédito. Nesse contexto é que se inserem os modelos quantitativos de risco de crédito para auxiliarem na concessão do crédito e administração do risco nas instituições de microcrédito, conferindo, possivelmente, maior eficiência operacional a essas instituições no atendimento de seus objetivos.

O objetivo do presente trabalho consistiu em analisar a possibilidade de utilização de modelos quantitativos de *Credit Scoring* na avaliação e prevenção do risco de inadimplência do Fundo Rotativo de Ação da Cidadania - Cred Cidadania.

A justificativa para a realização deste trabalho deve-se à importância do microcrédito no atual contexto econômico e financeiro do país, constituindo modalidade de crédito maciçamente utilizada por milhares de microempreendedores. Além disso, é importante mencionar que a realização deste trabalho se justifica também pelas possíveis contribuições que os seus resultados podem fornecer ao processo de concessão e análise do crédito no Cred Cidadania, bem como para outras instituições de microcrédito, todas muito carentes de instrumentos metodológicos de auxílio à gestão, principalmente do risco.

## **2. Revisão de literatura**

### **2.1 Modelos de *Credit Scoring***

Os modelos de *Credit Scoring* são sistemas que atribuem pontuações às variáveis de decisão de crédito de um proponente, mediante a aplicação de técnicas estatísticas. Esses modelos visam a segregação de características que permitam distinguir os bons dos maus créditos (LEWIS, 1992)

A partir de uma equação gerada através de variáveis referentes ao proponente de crédito e/ou à operação de crédito, os sistemas de *Credit Scoring* geram uma pontuação que representa o risco de perda. O score que resulta da equação de *Credit Scoring* pode ser interpretado como probabilidade de inadimplência ao se comparar a pontuação de um crédito qualquer com determinada pontuação estabelecida como ponto de corte ou pontuação mínima aceitável. Conforme ressalta Saunders (2000), o score pode ser utilizado para classificação

REAd – Edição 62 Vol 15 N° 1 jan-abril 2009

## CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

de créditos como adimplentes ou inadimplentes, bons ou maus, desejáveis ou não, de acordo com a pontuação obtida por cada crédito. Esta classificação, por sua vez, pode orientar a decisão do analista em relação à concessão ou não do crédito solicitado.

Assim, a idéia essencial dos modelos de *Credit Scoring* é identificar certos fatores-chave que influenciam na adimplência ou inadimplência dos clientes, permitindo a classificação dos mesmos em grupos distintos e, como consequência, a decisão sobre a aceitação ou não do crédito em análise.

Em relação aos usos do *Credit Scoring*, Caouette, Altman e Narayanan (1998) afirmam que embora estes sistemas sejam utilizados para decisões sobre a concessão ou não de crédito, que está centrada na avaliação do risco de crédito ou inadimplência, algumas instituições utilizam-no para determinação do tamanho do crédito a ser concedido. Marques (2002) acrescenta outras aplicações do *Credit Scoring*, como a sua utilização na classificação de créditos conforme o risco, e recentemente, são utilizados como base para o desenvolvimento de modelos mais complexos, a exemplo, aqueles baseados na teoria de carteiras.

Os modelos de *Credit Scoring* podem ser aplicados tanto à análise de crédito de pessoas físicas quanto empresas. Quando aplicados a pessoas físicas, eles utilizam informações cadastrais e de comportamento dos clientes. Já quando aplicados a empresas, são utilizados índices financeiros como variáveis determinantes ou não da insolvência das mesmas. Conforme ressalta Saunders (2000), a idéia é essencialmente a mesma: a pré-identificação de certos fatores-chave que determinam a probabilidade de inadimplência e sua combinação ou ponderação para produzir uma pontuação quantitativa.

Os modelos de *Credit Scoring* são divididos em duas categorias: modelos de aprovação de crédito e modelos de escoragem comportamental, também conhecidos por *Behavioural Scoring* (SAUNDERS, 2000).

Thomas (2000) explica as diferenças entre modelos de aprovação de crédito e modelos de escoragem comportamental. Segundo este autor, os modelos de *Credit Scoring* propriamente ditos são ferramentas que dão suporte à tomada de decisão sobre a concessão de crédito para novas aplicações ou novos clientes. Já os modelos *Behavioural Scoring* auxiliam na administração dos créditos já existentes, ou seja, aqueles clientes que já possuem uma relação creditícia com a instituição.

Desta forma, enquanto o principal objetivo dos modelos de aprovação de crédito é estimar a probabilidade de um novo solicitante de crédito se tornar inadimplente com a

instituição em determinado período, os modelos de escoragem comportamental objetivam estimar a probabilidade de inadimplência de um cliente que já possui um produto ou crédito com a instituição.

## **2.2 Algumas técnicas estatísticas aplicadas em modelos *Credit Scoring***

### **2.2.1 Análise discriminante**

A análise discriminante consiste em uma técnica estatística multivariada usada na resolução de problemas que envolvem separação de conjuntos distintos de objetos ou observações, e a alocação de novos objetos ou observações em conjuntos previamente definidos.

De acordo com Hair Júnior et al (1998), a análise discriminante implica a derivação de uma combinação linear de duas ou mais variáveis independentes que melhor discrimine ou diferencie grupos previamente definidos.

De uma forma geral, no contexto da aplicação da análise discriminante ao processo de avaliação de risco de crédito, admitindo a existência de dois grupos distintos, como clientes adimplentes e inadimplentes, ou empresas solventes e insolventes, deseja-se encontrar uma função, que nada mais é que uma combinação linear para melhor discriminar os grupos entre si. Ou seja, deseja-se obter um conjunto único de coeficientes para cada uma das variáveis independentes, que classifique, com precisão, cada elemento em um dos grupos previamente definidos.

De acordo com Hair Júnior et al (1998), a acima referida combinação linear para uma análise discriminante, também conhecida como função discriminante, provém de uma equação que tem a seguinte forma:

$$Z = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_i X_i$$

Onde:

Z = escores discriminantes (variável dependente);  $B_0$  = constante;  $B_i$  = coeficiente ou peso discriminante para a variável i;  $X_i$  = variável independente;

A partir da multiplicação de cada variável independente por seu coeficiente ou peso discriminante, obtêm-se os escores discriminantes para cada indivíduo. Esses escores são utilizados para classificar os indivíduos em cada grupo, o que é feito através da comparação com um escore de corte determinado pela média aritmética das médias dos escores de cada grupo (centróides de cada grupo).

### 2.2.2 Regressão logística

A análise *Logit* ou regressão logística consiste em uma técnica estatística utilizada na separação de dois grupos, que visa obter a probabilidade de que uma observação pertença a um conjunto determinado, em função do comportamento das variáveis independentes. (HAIR JÚNIOR ET AL, 1998). Ela é comumente utilizada para análise de dados com resposta binária ou dicotômica e consiste em relacionar, através de um modelo, a variável resposta (variável dependente binária) com fatores que influenciam ou não a probabilidade de ocorrência de determinado evento (variáveis independentes).

No âmbito da aplicação ao risco de crédito, a técnica de regressão logística é utilizada para a avaliação da inadimplência de determinado grupo de clientes em situações relativas à concessão de crédito, assumindo que a probabilidade de inadimplência é logisticamente distribuída, com resultado binomial 0 ou 1.

Supondo um evento dependente em que a variável  $Y$  é uma variável binária, que assume valores 0 ou 1; e variáveis independentes  $X_1, X_2, \dots, X_k$ , a função de distribuição logística, segundo Gujarati (2000), é dada por:

$$P_i = \Pr [Y_i = 1 | X = X_i] = \frac{1}{1 + e^{-Z}}, \text{ onde:}$$

$P_i$  Representa a probabilidade de um evento ocorrer;

$$Z = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_i X_i;$$

Analisando o significado da função de distribuição logística no contexto de risco de crédito, tem-se: a variável dependente consiste na situação de adimplência do cliente, que assumirá valores 0 ou 1, a depender dos dados procederem de um cliente adimplente ou inadimplente, ou, empresa solvente e insolvente, respectivamente; as variáveis independentes representam os fatores que se supõe influenciam a inadimplência, como, por exemplo, dados pessoais, econômicos, financeiros, dos clientes pessoas físicas; índices financeiros, no caso de empresas. A probabilidade de inadimplência do cliente ou de insolvência da empresa é dada por  $P_i$ , que é a probabilidade condicional de  $Y$  assumir o valor 1. Os coeficientes estimados  $\beta_s$  representam medidas das variações na proporção das probabilidades.

## 2.3 Microcrédito

Kwitko (1999) entende o microcrédito como um sistema de crédito diferenciado, destinado a setores com pouco ou nenhum acesso ao sistema financeiro tradicional, como os microempreendedores, inclusive os informais, os artesãos, os prestadores de serviços, os micronegócios familiares, as agroindústrias artesanais familiares, os autônomos, as microempresas formais, bem como as associações ou cooperativas constituídas pelos segmentos mencionados, visando a implantação, a modernização, a ampliação e/ou a diversificação de atividades capazes de gerar ou manter trabalho e renda, em bases auto-sustentáveis, mediante investimentos de pequeno valor, com respaldo principalmente no sistema de crédito solidário.

Desta forma, pode-se apresentar um resumo dos principais aspectos que caracterizam o microcrédito como um tipo diferenciado de crédito:

- a) Crédito produtivo: o microcrédito é um crédito especializado que objetiva apoiar pequenos empresários e microempreendedores que desejam investir no seu negócio. Está voltado para negócios de pequeno porte, gerenciados por pessoas de baixa renda, não se destinando, portanto, ao financiamento do consumo
- b) Sistema de garantias: A principal garantia utilizada no microcrédito é o aval solidário (ou fiança solidária), que consiste na reunião, em geral, de três a cinco pessoas com pequenos negócios e necessidades de crédito, que confiam umas nas outras para formar um grupo solidário, com o objetivo de assumir as responsabilidades pelos créditos de todo o grupo.
- c) Crédito orientado: A metodologia de concessão de crédito e avaliação dos riscos no microcrédito se caracteriza, principalmente, pelo acompanhamento dos créditos concedidos, o que é um aspecto de fundamental importância no microcrédito, já que orienta os tomadores para o emprego adequado dos recursos.
- d) Crédito adequado ao ciclo do negócio : As operações de microcrédito são consideradas adequadas ao ciclo de negócio, devido, principalmente, ao fato dos empréstimos serem baseados em valores pequenos inicialmente, mas crescentes e aos curtos prazos de pagamento, geralmente semanais, quinzenais ou, no máximo, mensais.

### 2.3.1 Análise de riscos em instituições de microcrédito

No microcrédito, a avaliação do risco de inadimplência está sedimentada na análise de crédito tradicional, caracterizada por elementos qualitativos, porém especialmente adaptada a essas instituições. Segundo Kwitko (1999), os principais aspectos considerados na análise de risco na concessão de microcrédito dizem respeito aos C's do crédito (Caráter, Capacidade, REAd – Edição 62 Vol 15 N° 1 jan-abril 2009

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

Capital, Colateral, Condições), que, no entanto, nas instituições de microcrédito, buscam identificar características do empreendedor e do seu negócio. Portanto, os C's do crédito constituem os fatores de risco a serem considerados quando da análise de risco de inadimplência nas instituições de microcrédito, sendo a decisão sobre a concessão ou renovação de um crédito centrada na avaliação qualitativa desses fatores.

Segundo Bruett (2002), a despeito das dificuldades em se utilizar metodologias quantitativas na avaliação de risco em instituições de microcrédito, alguns trabalhos dessa natureza vêm sendo realizados atualmente. Pode-se citar, como exemplo, o trabalho de Schreiner (1999) em que o autor desenvolveu um modelo de *Credit Scoring* para uma instituição de microcrédito na Bolívia.

### 3 Metodologia

Foram desenvolvidos, neste trabalho, dois tipos de modelos de *Credit Scoring* distintos: modelos de aprovação de crédito (*Credit Scoring* propriamente dito) e modelos de escoragem comportamental (*Behavioural Scoring*).

Os modelos de aprovação de crédito utilizaram informações constantes do cadastro do cliente e do seu negócio. Esses modelos têm como objetivo principal servirem de ferramentas de auxílio à avaliação e decisão do analista sobre a concessão ou não de crédito a um novo cliente.

Nos modelos de escoragem comportamental (*Behavioural Scoring*) foram incorporadas variáveis que retratam a história do cliente com a instituição. Esses modelos objetivam auxiliar o analista de crédito em suas decisões sobre renovações de empréstimos de clientes, renegociações de dívidas, determinação de montante para renovação de empréstimo, determinação do valor da parcela em renovação de empréstimo, enfim, todas as decisões relativas ao gerenciamento do crédito de clientes que já possuem uma relação ou um histórico com a instituição.

Para construção dos modelos, tanto os de aprovação quanto aqueles *behavioural scoring*, foram empregadas as técnicas estatísticas análise discriminante e regressão logística, que são, atualmente, amplamente utilizadas para construção de modelos *Credit Scoring* e que foram devidamente explanados em seção anterior deste trabalho.



### 3.1 Amostra e coleta de dados

A amostra utilizada neste trabalho foi composta por 200 clientes sendo 100 adimplentes e 100 inadimplentes. Este tamanho amostral se mostrou estatisticamente relevante ao nível de confiança de 90%, para a população de 1400 clientes do Cred Cidadania. Os dados foram coletados a partir do sistema computacional utilizado pela instituição, denominado SISGEM (Sistema de Gestão de Microcrédito).

Os dados coletados referem-se a operações de créditos concedidas a clientes do Cred Cidadania no período de outubro a dezembro de 2004. Todos os créditos concedidos nesse período foram analisados com intuito de verificar a condição de adimplência de cada um no mês de julho de 2005. A escolha do mês de julho como período de referência para avaliação da performance do cliente deveu-se ao fato de que o número de parcelas dos créditos concedidos pela instituição é, em média, 5. Sendo assim, até julho de 2005, tem-se um horizonte temporal suficiente para avaliar se os créditos concedidos em outubro, novembro e dezembro de 2004 estavam adimplentes ou não.

Os clientes selecionados através do processo de amostragem aleatória tiveram suas fichas cadastrais e do negócio impressas. A partir desses registros foram identificadas as características pessoais dos clientes e dados econômico-financeiros relativos ao negócio ou atividade econômica dos mesmos, tanto dos adimplentes quanto dos inadimplentes.

### 3.2 Definição das variáveis

A variável resposta (dependente) nos modelos de *Credit Scoring* é a qualidade de crédito (adimplência ou inadimplência) da operação de empréstimo. Neste trabalho, foram denominados clientes inadimplentes aqueles clientes com atraso superior a 60 dias em pelo menos uma parcela do empréstimo. Em relação aos adimplentes, foram considerados integrantes deste grupo aqueles clientes que não possuíam atrasos ou que possuíam atrasos de no máximo 30 dias em alguma parcela do empréstimo. Aqueles clientes com atraso superior a 30 dias e inferior a 60 dias foram considerados indefinidos e foram eliminados da amostra.

Para classificar as observações de acordo com a qualidade de crédito, foram selecionadas variáveis explicativas ou independentes que pudessem influenciar a situação de adimplência dos clientes em suas operações de empréstimos. A identificação inicial das variáveis explicativas foi baseada em estudos anteriores sobre o assunto e em informações obtidas a partir do convívio no Cred Cidadania.

As variáveis iniciais pré-selecionadas foram: Estado Civil do Cliente, Gênero do Cliente, Natureza da atividade econômica do negócio, Local de Residência do Cliente, Idade

REAd – Edição 62 Vol 15 N° 1 jan-abril 2009

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

do Cliente, Número de Dependentes Financeiros do Cliente, Receita Familiar Bruta, Despesa Familiar Bruta, Resultado Líquido Familiar, Receita Bruta do Negócio, Despesa Bruta do Negócio, Resultado Líquido do Negócio, Resultado Líquido Total, Tempo de Funcionamento do Negócio, Receita Bruta do Avalista, Despesa Bruta do Avalista, Resultado Líquido do Avalista, Valor do Último Empréstimo, Valor da Parcela do Último Empréstimo, Número de Parcelas do Último Empréstimo, Percentual de Endividamento, Número de Créditos Anteriores com a Instituição, Histórico do Titular com a Instituição, Histórico do Avalista com a Instituição, Agente de Crédito Responsável pelo Empréstimo.

As variáveis explicativas de natureza qualitativa foram inseridas na base de dados através de variáveis *dummy*, que consistem em variáveis construídas artificialmente para mensurar a presença ou ausência de algum atributo de natureza qualitativa. Elas assumem valor 0 ou 1, dependendo, respectivamente, da ausência ou presença de determinado atributo. Assim, foram inseridas  $N - 1$  *dummies* para representar as  $N$  categorias de cada variável qualitativa anteriormente explanada.

### 3.3 Preparação dos dados, estimação e avaliação dos modelos

A fase inicial da preparação dos dados consistiu em estruturar uma base agregando os valores numéricos referentes ao conjunto das possíveis variáveis explicativas pré-selecionadas para utilização na construção dos modelos.

Depois de estruturada a base de dados, as variáveis independentes quantitativas foram testadas quanto à normalidade de suas distribuições, que é um pressuposto adotado para modelagem adotando a técnica estatística análise discriminante. O teste de normalidade empregado foi o Kolmogorov Smirnov  $Z$ . Para aquelas variáveis que o teste de Kolmogorov Smirnov acusou rejeição da hipótese de normalidade, foi realizada uma transformação, através do logaritmo de seus valores, com o intuito de tornar as distribuições semelhantes à normal.

A estimação dos modelos de *Credit Scoring* foi realizada através do método *stepwise*, que envolve a inclusão das variáveis independentes na função discriminante ou na regressão logística, uma por vez, com base em seu poder discriminatório. Após o desenvolvimento dos modelos de *Credit Scoring*, avaliou-se a capacidade preditiva de cada um deles com intuito de verificar o grau de ajustamento das funções construídas. Como ressalta Hair Júnior et al (1998), devido ao caráter não métrico da variável dependente, não se pode usar uma medida

como  $R^2$  para avaliar a capacidade preditiva dos modelos. Assim, em trabalhos de *Credit Scoring*, é comum a construção de matrizes de classificação para este fim.

As matrizes de classificação consistem em tabelas que comparam a classificação realizada pelos modelos desenvolvidos com a classificação original das observações da amostra. Essas matrizes são elaboradas através da análise de cada observação, visando conhecer se ela foi corretamente classificada pelos modelos. Os resultados dessa análise são os percentuais de acerto e erro de classificação dos modelos. Deste modo, as matrizes de classificação avaliam grau de ajuste dos modelos através da mensuração da eficiência dos mesmos em classificar corretamente os clientes da amostra.

O nível de significância da precisão de classificação dos modelos foi avaliado através do Teste Q de Press. Segundo Hair Júnior et al (1998), esse teste avalia o poder discriminatório da matriz de classificação do modelo quando comparada à classificação pelo critério de chances.

$$Q_{Pr\ ess} = \frac{[N - (n \times K)]^2}{N \times (K - 1)}, \text{ onde:}$$

N = Tamanho da amostra total;

n = Número de observações corretamente classificadas;

K = Número de grupos;

## 4 Resultados

### 4.1 Modelo de aprovação de crédito – técnica análise discriminante

O modelo final de aprovação de crédito, utilizando a técnica análise discriminante, bem como os parâmetros de classificação do modelo são dados nos quadros 1 e 2

Quadro 1: Modelo de aprovação de crédito – análise discriminante

Variáveis / Siglas	Coefficientes Estimados
Resultado Líquido do Negócio (RLN)	1,377
Número de Parcelas (NP)	0,805
Valor do Empréstimo (VE)	-1,454
Tempo de Funcionamento (TF)	-0,477
Constante ( $\beta_0$ )	-3,146

Fonte: Dados da Pesquisa

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

Quadro 2 : Médias dos escores e parâmetros de classificação do modelo de aprovação de crédito - análise discriminante

Escores	Cientes Adimplentes	Cientes Inadimplentes
Média (Centróides)	-0,665	0,665
Parâmetro de Classificação (ponto de corte)	Escore < 0	Escore > 0

Fonte: Dados da Pesquisa

Segundo Hair Jr. et (1998), o coeficiente Lambda de Wilks é o mais comumente utilizado para análise de significância de modelos construídos com análise discriminante. Esse coeficiente examina se os grupos considerados na variável dependente são, de algum modo, diferentes. Quanto maior a dispersão entre os grupos considerados, menor o Lambda de Wilks e maior a sua significância.

Quadro 3: Teste significância para o modelo de aprovação de crédito

Valor do Teste Lambda de Wilks	Significância
0,691	0,000

Fonte: Dados da Pesquisa

Verifica-se que o teste Lambda de Wilks é altamente significativo a qualquer nível de confiança (90%, 95% ou 99%). Isso indica que há diferenças entre os dois grupos de clientes (adimplentes e inadimplentes) e que o poder de separação da função discriminante é estatisticamente significativa.

O efeito de cada variável explicativa do modelo sobre os escores da operação de crédito pode ser descrito através da análise dos coeficientes:

- Resultado líquido do negócio: O sinal positivo do coeficiente dessa variável significa que clientes com maiores rendas líquidas possuem maior tendência a serem mais inadimplentes. Esse resultado, a princípio incoerente, possivelmente, tem como causa o fato dos clientes declararem aos agentes de crédito renda superior àquela realmente auferida. Uma vez que a maioria dos empreendimentos financiada pela instituição é informal, não é exigida comprovação formal de renda. O solicitante declara quais são suas receitas provenientes do negócio e o agente de crédito faz uma avaliação do empreendimento para verificar se as condições do negócio condizem com a renda declarada. Assim, o coeficiente desta variável mostra que pode estar havendo uma distorção de informações dos clientes, que declaram receitas maiores que as realmente auferidas, apenas para conseguirem o crédito.

- Número de parcelas: empréstimos com pagamentos divididos em maiores números de parcelas tendem a ser mais inadimplentes. Isso, provavelmente, está ligado à finalidade do empréstimo. Os empréstimos destinados a investimento em capital fixo possuem prazos maiores para pagamento que aqueles destinados a capital de giro. No entanto, os possíveis retornos para o empreendimento, gerados com o investimento em capital fixo, acontecem de forma mais lenta, às vezes, posteriormente ao prazo final de pagamento do empréstimo. Portanto há uma tendência de clientes com empréstimos destinados a capital fixo possuírem maiores dificuldades para pagamento, principalmente, nas primeiras parcelas do empréstimo, sendo, assim, caracterizados como mais arriscados.
- Valor do empréstimo: de acordo com esse modelo, empréstimos em maiores valores tendem a ser mais adimplentes que empréstimos menores. Esse resultado está ligado à metodologia de concessão de crédito adotada no Cred Cidadania, em que os créditos são concedidos em valores crescentes, havendo aumento dos valores dos empréstimos de acordo com a capacidade de pagamento e pontualidade do usuário. Deste modo, geralmente, a instituição concede créditos em maiores valores para aqueles clientes que já possuem um histórico de bons pagamentos com a mesma.
- Tempo de funcionamento: empréstimos para empreendimentos com maior horizonte temporal de funcionamento são mais adimplentes que aqueles empréstimos direcionados a empreendimentos mais novos. Geralmente, um maior tempo de funcionamento do negócio é um indício de que aquele cliente já está mais estabilizado financeiramente e no mercado em que atua, o que é coerente com os resultados encontrados.

Além da direção do efeito das variáveis explicativas sobre os escores, é importante analisar também a magnitude ou tamanho do efeito dessas variáveis sobre os mesmos. A análise do quadro 1 mostra que, neste modelo de aprovação de crédito – análise discriminante, o valor do empréstimo (VE) é a característica que exerce o maior efeito na redução da inadimplência. Por outro lado, a variável que mais afeta a inadimplência, no sentido de aumentá-la, é o resultado líquido do negócio (RLN), já que seu coeficiente é o maior. Assim, por exemplo, ao realizar uma classificação de um solicitante com base no modelo construído, o fato dele declarar altos valores (muito superiores à média) para a variável resultado líquido do negócio irá contribuir significativamente para sua classificação como possível cliente inadimplente.

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

## 4. 2 Modelo *behavioural scoring* – técnica análise discriminante

Além das variáveis consideradas no modelo de aprovação de crédito, o modelo final *behavioural scoring* incorporou também variáveis que retratam o histórico dos clientes com a instituição, adicionando informações que auxiliem o analista a decidir sobre gestão de créditos de solicitantes que já são clientes da instituição. O modelo final de *behavioural scoring*, construído com o uso da técnica análise discriminante, pode ser visualizado no quadro 4 a seguir:

Quadro 4 - Modelo *behavioural scoring* – análise discriminante

Variáveis / Siglas	Coefficientes Estimados
Resultado Líquido do Negócio (RLN)	1,100
Número de Parcelas (NP)	0,678
Valor do Empréstimo (VE)	-0,702
Tempo de Funcionamento (TF)	-0,427
Não casados (EC <sub>1</sub> )	0,537
Número de Dependentes (ND)	0,216
Agente de Crédito 5 (AC <sub>5</sub> )	-0,998
Cliente não possui atrasos anteriores (CNPA)	-0,672
Constante ( $\beta_0$ )	-6,017

Fonte: Dados da Pesquisa

Quadro 5: Médias dos escores e parâmetros de classificação do modelo *behavioural scoring* – análise discriminante

Escores	Cientes Adimplentes	Cientes Inadimplentes
Média (Centróides)	-0,813	0,813
Parâmetro de Classificação (Ponto de Corte)	Escore < 0	Escore > 0

Fonte: Dados da Pesquisa

A significância do modelo *behavioural scoring*, dada pelo teste Lambda de Wilks, pode ser visualizada no quadro 6 abaixo:

Quadro 6: Teste de Significância para o modelo *behavioural scoring* – análise discriminante

Valor do Teste Lambda de Wilks	Significância
0,600	0,000

Fonte: Dados da Pesquisa

Uma vez que o valor do teste Lambda de Wilks foi baixo (consideravelmente inferior a 1) e significativo em qualquer nível de confiança, pode-se dizer que a função possui poder de discriminação estatisticamente relevante.

Os coeficientes das variáveis que já foram consideradas no modelo de aprovação (RLN, NP, VE, TF) apresentaram os mesmos sinais, e, portanto, efeito similar sobre os escores de cada cliente, o que indica não haver incoerência nessas variáveis entre os dois modelos. Além dessas variáveis já consideradas no modelo de aprovação – análise discriminante, os modelos *behavioural scoring* incorporaram outras variáveis, cujos efeitos sobre os escores são analisados a seguir.

- Estado civil não casado (solteiros, viúvos, divorciados): esta variável possui sinal positivo, significando que indivíduos não casados são mais propensos à inadimplência que indivíduos casados. Uma possível explicação para o comportamento dessa variável reside no fato dos indivíduos não casados possuírem uma menor apreensão em relação ao pagamento das dívidas, já que, geralmente, têm menores responsabilidades.
- Número de dependentes: tomadores com maior número de dependentes tendem a ser mais inadimplentes. Esse é um resultado plausível, uma vez que um maior número de dependentes, geralmente, significa que o cliente tomador de crédito compromete um maior percentual da renda familiar com o provimento das necessidades dos seus dependentes, visto que a família é mais numerosa. Além disso, como a renda familiar e do negócio são, na maioria dos casos, associadas, as receitas provenientes do negócio costumam também ser destinadas ao suprimento de necessidades da família do cliente.
- Agente de crédito 5 : o coeficiente dessa variável é negativo, indicando que a atuação desse agente de crédito contribui para a redução da inadimplência. Em outras palavras, os créditos monitorados pelo agente de crédito 5 são menos propensos a ficarem inadimplentes, o que denota a eficiência desse profissional, comparada aos outros agentes, na análise e acompanhamento dos empréstimos. O perfil desse agente é o fator decisivo na sua eficiência. Primeiramente, ele é o agente de crédito mais antigo da instituição, possui grande experiência em lidar com os clientes, na avaliação do risco e nas atividades de cobrança. Além da experiência, o agente de crédito 5 também é o mais motivado e participante em todas as atividades de capacitação profissional desenvolvidas pela instituição. O coeficiente dessa variável mostra a importância da experiência e capacitação dos agentes de crédito, visto que a atuação destes profissionais é peça fundamental no processo de análise de risco do Cred Cidadania e da maioria das instituições de microcrédito.

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

- Cliente não possui atraso anterior: o fato do solicitante do crédito possuir um histórico de bons pagamentos com a instituição é algo positivo, sendo que clientes nessa condição são menos propensos a se tornarem inadimplentes. O conhecimento, por parte da instituição, do histórico de atrasos dos clientes, é de fundamental importância, fornecendo subsídio de informação em decisões de renovação de empréstimos, aumento do valor concedido, dentre outras.

Em relação à magnitude do efeito dos coeficientes estimados sobre a inadimplência, verifica-se que, neste modelo *behavioural scoring* – análise discriminante, a variável que mais contribui para o aumento da inadimplência é o resultado líquido do negócio (RLN), sendo que quanto maiores forem os valores dessa variável, maior será a probabilidade do cliente ser ou se tornar inadimplente. Tal como no modelo anterior de aprovação de crédito, o coeficiente dessa variável evidencia a necessidade de avaliação mais pormenorizada para aqueles solicitantes que declarem rendas líquidas muito superiores à média. Por outro lado, a variável agente de crédito 5 (AC<sub>5</sub>) possui um coeficiente de -0,998, constituindo, portanto, a característica que exerce o maior impacto na redução da inadimplência, dentre as variáveis consideradas pelo modelo, devido à já referida eficiência deste profissional.

## 4.3 Modelo de aprovação de crédito – técnica regressão logística

O uso da técnica regressão logística gerou modelos próximos àqueles gerados com o uso da análise discriminante. O modelo de aprovação de crédito teve a seguinte composição, como mostra o quadro 7.

Quadro 7: Modelo de aprovação de crédito – regressão logística

Variáveis / Siglas	Coefficientes Estimados	Valor Teste Wald	Significância do Teste
Resultado Líquido do Negócio (RLN)	0,01	29,402	0,000
Número de Parcelas (NP)	1,530	12,052	0,001
Valor do Empréstimo (VE)	-0,020	24,677	0,000
Tempo de Funcionamento (TF)	-0,083	5,894	0,015
Local de Residência 4 (LOCAL4)	-1,246	7,093	0,008
Constante ( $\beta_0$ )	-7,032	10,396	0,001

Fonte: Dados da Pesquisa



Verifica-se que esse modelo de aprovação de crédito com regressão logística agregou apenas uma variável a mais em relação ao modelo construído com análise discriminante, mostrado na seção 4.1 deste trabalho. Embora os valores dos coeficientes estimados para as variáveis explicativas sejam diferentes, todos eles possuem os mesmos sinais que os coeficientes gerados no modelo de aprovação com análise discriminante, indicando que a influência das variáveis explicativas sobre os escores é similar nos dois modelos. Desta forma, uma variável com coeficiente positivo afeta o escore aumentando a tendência à inadimplência enquanto que um coeficiente positivo a reduz.

Enquanto na análise discriminante as previsões de pertinência ou classificação dos indivíduos em cada um dos grupos (adimplentes ou inadimplentes) são realizadas através dos escores médios dos grupos, a regressão logística classifica os clientes prevendo diretamente a probabilidade do evento inadimplência acontecer. Lembrando que a variável dependente  $Y_i$  representa a qualidade de crédito do indivíduo  $i$ , a regressão logística calcula diretamente a probabilidade condicional de  $Y_i$  ser igual a 1, que, nesse estudo, significa a probabilidade do cliente ser inadimplente. A probabilidade condicional de  $Y_i$  ser igual a 1, dadas as respostas das variáveis explicativas do indivíduo  $i$ , é fornecida pela seguinte fórmula :

$$P_i = \frac{1}{1 + e^{-Z_i}} ,$$

Onde:

$Z_i$  = é a equação representativa do modelo de aprovação de crédito estimado;

$P_i$  = representa a probabilidade do indivíduo  $i$  estar inadimplente;

Desta forma, a classificação dos clientes como adimplentes ou inadimplentes, neste modelo, foi realizada com base na probabilidade de inadimplência, que é calculada de acordo com equação gerada pelo modelo, que, por sua vez, se baseia nas informações de cada cliente para as variáveis explicativas. O ponto de corte adotado foi 0,5, valor padronizado para a técnica regressão logística. Esse valor de 0,5, conforme ressalta Hair Júnior et al (1998), representa a probabilidade de ocorrência do evento segundo o critério de aleatoriedade ou chances iguais. Assim, aqueles clientes para os quais a probabilidade estimada de inadimplência resultou em valor inferior a 0,5, foram classificados como adimplentes. E aqueles para os quais a probabilidade de inadimplência foi superior a 0,5, foram classificados como inadimplentes.

## CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

Em relação às variáveis explicativas do modelo de aprovação, além das variáveis explicativas já consideradas no modelo de aprovação com análise discriminante, cujos efeitos sobre a variável dependente já foram explanados na abordagem dos modelos de análise discriminante, o modelo de aprovação - regressão logística incorporou também a variável local de residência 4, que será comentada a seguir:

- **Local de Residência 4:** Essa variável agrupa clientes residentes nas localidades de Paulista, Jaboatão, Igarassu, Itamaracá, Abreu e Lima, Moreno e São Lourenço. O sinal negativo do coeficiente estimado indica que os clientes residentes nessas localidades possuem menores chances de se tornarem inadimplentes do que clientes residentes nas outras localidades (Camaragibe, Recife, Olinda). Há indícios, de acordo com levantamentos realizados na instituição, de que tal resultado reflete a atuação dos agentes de crédito em cada região geográfica.

Na regressão logística, é utilizada a estatística de Wald para testar a significância da equação estimada (HAIR JÚNIOR ET AL, 1998). Ela fornece a significância estatística de cada coeficiente estimado, de modo que o teste de hipóteses pode ocorrer como acontece na regressão múltipla. Os valores dos testes de Wald para os coeficientes do modelo de aprovação de crédito foram dados no quadro 7. Verifica-se que, a um nível de 95% de confiança, todos os coeficientes foram estatisticamente significativos ou diferentes de 0, já que a significância do teste é inferior a 0,05 para todos eles. Assim, uma vez que todos os coeficientes foram estatisticamente significativos, considerou-se que a regressão, em sua forma global, também fora significativa.

#### **4.4 Modelo *behavioural scoring* – técnica regressão logística**

O modelo *behavioural scoring* com regressão logística adicionou duas variáveis a mais que o modelo de aprovação de crédito com a mesma técnica. Em comparação com o modelo *behavioural scoring* elaborado com análise discriminante, o modelo *behavioural scoring* com regressão logística se diferenciou em relação a três variáveis. O modelo *behavioural scoring* com regressão logística é mostrado no quadro 8:

Quadro 8: Modelo *behavioural scoring* – regressão logística

Variáveis / Siglas	Coefficientes Estimados	Valor Teste Wald	Significância do Teste
Resultado Líquido do Negócio (RLN)	0,01	28,740	0,000
Número de Parcelas (NP)	1,538	11,364	0,001
Valor do Empréstimo (VE)	-0,02	13,643	0,000
Tempo de Funcionamento (TF)	-0,099	6,822	0,009
Número de Dependentes (ND)	0,312	4,731	0,03
Agente de Crédito 5 (AC <sub>5</sub> )	-1,185	5,270	0,022
Local de Residência 4 (LOCAL 4)	-1,183	5,678	0,017
Constante ( $\beta_0$ )	-6,017	11,570	0,001

Fonte: Dados da Pesquisa

Verifica-se, através da análise do quadro, que os coeficientes estimados no modelo *behavioural scoring* – regressão logística influenciam a variável dependente qualidade do crédito no mesmo sentido que os coeficientes estimados no modelo *behavioural scoring* – análise discriminante. Assim, as variáveis explicativas cujos coeficientes são positivos contribuem para um aumento da inadimplência enquanto que as variáveis com sinais negativos reduzem-na. Todas as variáveis que formaram o modelo *behavioural scoring* – regressão logística já foram comentadas nos modelos anteriores, e como elas afetam os escores na mesma direção, os comentários são válidos também para este modelo.

De modo idêntico ao modelo *behavioural scoring* – análise discriminante, neste modelo com regressão logística, o coeficiente que mais afeta os escores na direção de redução da inadimplência é o coeficiente da variável agente de crédito 5 (AC5). Já o fato do empréstimo possuir amplo número de parcelas (variável NP) é o que mais colabora para aumentar a propensão à inadimplência. Comentários sobre essas variáveis já foram feitos anteriormente, quando da consideração de outros modelos. Mas vale ressaltar que o fato da variável agente de crédito 5 ser a que mais impacta na redução da inadimplência vem confirmar, também neste modelo com regressão logística, a relevância dos aspectos experiência e formação do agente para que ele exerça com eficácia suas funções, contribuindo para a redução da inadimplência do Cred Cidadania.

Em relação à significância estatística do modelo, o teste Wald mostrou que todos os coeficientes são significativos a um nível de 95%, uma vez que o valor da significância do teste foi inferior a 0,05 para todos eles. Assim, comprovou-se que o modelo, em sua forma geral, foi significativo.

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

Por fim, é importante mencionar que, em todos os modelos, as constantes ( $\beta_0$ ) figuram apenas para fins de arranjos estatísticos, sendo que estas constantes não são consideradas para fins interpretativos, uma vez que os modelos são padronizados pelo pacote estatístico para trabalhar apenas com os coeficientes das variáveis independentes.

## 4.5 Avaliação da capacidade preditiva dos modelos

### 4.5.1 Matrizes de classificação dos modelos de análise discriminante

Quadro 9 - Matriz de classificação do modelo de aprovação de crédito – análise discriminante (% acertos)

Classificação Original	Classificação do Modelo	
	Adimplentes	Inadimplentes
Adimplentes	84%	16%
Inadimplentes	23%	77%
Percentual Geral de Acertos	80,5%	

Fonte: Dados da Pesquisa

Quadro 10 - Matriz de classificação do modelo de *behavioural scoring* – análise discriminante (% acertos)

Classificação Original	Classificação do Modelo	
	Adimplentes	Inadimplentes
Adimplentes	84%	16%
Inadimplentes	21%	79%
Percentual Geral de Acertos	81,5%	

Fonte: Dados da Pesquisa

Verifica-se, através da análise dos quadros 09 e 10, que, de uma forma geral, os modelos desenvolvidos com o uso da técnica análise discriminante obtiveram bons percentuais de acertos em suas classificações, situando-se na faixa de 80% de acertos. O Teste Q de Press foi empregado para avaliar o nível de significância da precisão de classificação, comparando a classificação correta realizada pelos modelos com aquela baseada no critério de chances, que para o caso de dois grupos é de 50%.

$$Q_{Pr ess} = \frac{[N - (n \times K)]^2}{N \times (K - 1)}$$

Considerando o número de observações corretamente classificadas em cada modelo, o tamanho total da amostra  $N = 200$  e o número de grupos  $n=2$ , a aplicação da fórmula acima mostrada para cada um dos modelos gerou os seguintes resultados:

- Modelo de aprovação de crédito – análise discriminante

$$Q_{\text{Press}} = \frac{[200 - (161 \times 2)]^2}{200(2 - 1)} = 74,42$$

- Modelo *Behavioural Scoring* – análise discriminante

$$Q_{\text{Press}} = \frac{[200 - (163 \times 2)]^2}{200(2 - 1)} = 79,38$$

Os valores de Q de Press encontrados foram comparados com o valor crítico ou tabelado para esse teste, que é o valor de  $\chi^2$  (qui-quadrado) com um grau de liberdade, já que o número de graus de liberdade consiste no número de grupos considerados menos 1, ou seja,  $(K - 1)$ . Adotando-se o nível de confiança de 95%, o valor de  $\chi^2$ , com um grau de liberdade é 3,84. Uma vez que os valores dos testes Q de Press para ambos modelos de análise discriminante excedem, em muito, o valor de  $\chi^2$  crítico ou tabelado, constatou-se que as matrizes de classificação dos dois modelos podem ser consideradas estatisticamente melhores que a classificação baseada no critério de chances.

Segundo Hair Júnior et al (1998), não existe nenhuma regra que defina qual deve ser a precisão da capacidade preditiva dos modelos. O autor afirma que isso depende da finalidade do modelo, do pesquisador e do custo da informação gerada em relação ao seu valor. No entanto, segundo ele, uma estimativa grosseira do nível aceitável de precisão preditiva dos modelos pode ser obtida considerando que a precisão da classificação através do modelo deve ser, pelo menos, um quarto superior à classificação por chances. Assim, se a classificação por chances é 50%, a classificação do modelo deve ser, pelo menos, 62,5%. Os modelos de análise discriminante desenvolvidos neste trabalho resultaram em níveis de acerto de 80,5% e 81,5%, obtendo, portanto uma precisão de classificação consideravelmente superior ao mínimo aceitável.

Embora os modelos desenvolvidos com análise discriminante tenham obtido bons percentuais gerais de acerto, verifica-se, ao analisar os quadros 10 e 11, que a sua precisão em classificar os clientes inadimplentes foi inferior à precisão de classificação dos clientes adimplentes. Para uma instituição concessora de crédito, é mais interessante que o modelo seja eficaz na previsão da inadimplência do que da adimplência, já que o erro de aprovar uma operação que se tornará problemática (erro tipo I) é considerado mais grave que a recusa de

REAd – Edição 62 Vol 15 N° 1 jan-abril 2009

# CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

uma operação que seria um bom negócio para a instituição (erro Tipo II), como ressalta Sicsu (1998).

Assim, pode-se dizer que ambos modelos de análise discriminante, apesar de terem sido eficazes na classificação geral, apresentaram-se menos poderosos em prever o evento mais necessário de ser previsto por um modelo de risco de crédito, qual seja, a inadimplência.

## 4.5.2 Matrizes de classificação dos modelos de regressão logística

Quadro 11 - Matriz de classificação do modelo de aprovação de crédito – regressão logística  
(% acertos)

Classificação Original	Classificação do Modelo	
	Adimplentes	Inadimplentes
Adimplentes	78%	22%
Inadimplentes	18%	82%
Percentual Geral de Acertos	80,0%	

Fonte: Dados da Pesquisa

Quadro 12 - Matriz de classificação do modelo *behavioural scoring* – regressão logística  
(% acertos)

Classificação Original	Classificação do Modelo	
	Adimplentes	Inadimplentes
Adimplentes	79%	21%
Inadimplentes	17%	83%
Percentual Geral de Acertos	81,0%	

Fonte: Dados da Pesquisa

O teste Q de Press para cada um dos modelos de regressão logística resultou nos seguintes valores:

- Modelo de aprovação de crédito – regressão logística

$$Q_{Pr\ ess} = \frac{[200 - (160 \times 2)]^2}{200(2 - 1)} = 72,00$$

- Modelo *behavioural scoring* – regressão logística

$$Q_{Pr\ ess} = \frac{[200 - (162 \times 2)]^2}{200(2 - 1)} = 76,88$$

De modo similar aos modelos de análise discriminante, com base no teste Q de Press, verificou-se que as classificações dos modelos de regressão logística foram muito superiores  
REAd – Edição 62 Vol 15 N° 1 jan-abril 2009

que a classificação com base no critério de chances. Os percentuais de acertos de ambos os modelos superam, com folga, o percentual mínimo aceitável aconselhado por Hair Júnior et al (1998) que é de 62,5%, ou um quarto superior à classificação por chances, que para o caso de dois grupos é de 50%. Isso significa que esses modelos possuem poder discriminatório de classificação dos grupos de clientes adimplentes e inadimplentes superior a uma classificação baseada apenas na probabilidade ou critério de chances.

Os percentuais gerais de acerto dos modelos de regressão logística foram muito próximos àqueles obtidos através da análise discriminante, conforme pode ser visto pela análise dos quadros 09 a 12. No entanto, verifica-se que os modelos de regressão logística foram relativamente mais eficazes em classificar corretamente os clientes inadimplentes. Enquanto os modelos de aprovação – análise discriminante classificaram corretamente 77% dos clientes inadimplentes, o percentual de classificação desses clientes de acordo com o modelo de aprovação – regressão logística foi de 82,0%. Resultados semelhantes foram encontrados em relação aos modelos behavioural scoring, em que a análise discriminante classificou corretamente 79% dos clientes inadimplentes contra 82% de classificação correta da técnica de regressão logística. Desta forma, sob o aspecto da previsão de inadimplência, os modelos de regressão logística foram superiores àqueles desenvolvidos com análise discriminante, embora os percentuais gerais de acertos da classificação dos modelos sejam muito próximos com as duas técnicas.

Referindo-se às classificações de modelos de análise discriminante e regressão logística, Hair Júnior et al (1998) consideram que quando as suposições básicas são atendidas, ambas técnicas oferecem resultados classificatórios e preditivos similares. Realmente, neste trabalho, os resultados mostram que as duas técnicas possuem capacidades preditivas similares, conforme foi analisado acima. No entanto, embora tenham sido realizadas todas as transformações para atender aos pressupostos básicos da análise discriminante, essa técnica se mostrou menos competente em classificar corretamente os clientes inadimplentes, o que é o principal benefício de um modelo de previsão de risco de crédito. Os modelos elaborados com a técnica de regressão logística se mostraram mais robustos e eficientes para esta função.

#### **4.6 Validação dos Resultados**

O estágio final na construção de modelos de *Credit Scoring* envolve a validação dos resultados. Diante da impossibilidade de se trabalhar com uma amostra de teste, foi adotado, neste trabalho um procedimento de validação alternativo, abordado em Hair Júnior et al (1998). Segundo os autores, uma técnica alternativa de validação consiste em estabelecer o

## CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

perfil dos grupos sobre as variáveis independentes para garantir sua correspondência com as bases conceituais do modelo desenvolvido. Esse perfil é traçado a partir das características dos clientes, analisando os valores médios de cada variável independente para ambos os grupos. Para se obter os perfis dos grupos de clientes adimplentes e inadimplentes, foram computadas, para cada grupo, as médias das variáveis independentes consideradas pelos modelos e foi verificada a coerência desses resultados com o comportamento das variáveis no modelo, em termos da forma como elas impactam na qualidade de crédito dos clientes. Esses dados são mostrados no quadro 13.

Quadro 13: Médias das variáveis explicativas para os grupos de clientes adimplentes e inadimplentes

Variáveis / Siglas	Médias dos Grupos		Teste T diferença de Médias	Significância do Teste T
	Adimplentes	Inadimplentes		
Resultado Líquido do Negócio (RLN)	1455,63	2971,55	4,699	0,000
Número de Parcelas (NP)	4,88	5,48	2,254	0,025
Valor do Empréstimo (VE)	1338,10	1181,90	0,482	0,630
Tempo de Funcionamento (TF)	9,25	6,91	2,832	0,005
Não casados (EC <sub>1</sub> )	0,32	0,45	1,897	0,059
Número de Dependentes (ND)	0,79	1,23	2,300	0,022
Agente de Crédito 5 (AC <sub>5</sub> )	0,31	0,13	3,132	0,002
Cliente não possui atrasos anteriores	0,43	0,24	2,891	0,004
Local de Residência 4 (LOCAL4)	0,25	0,16	2,981	0,003

Fonte: Dados da Pesquisa

A análise das variáveis explicativas do modelo nos dois grupos de clientes mostra o perfil desses grupos em relação à inadimplência. As tendências de comportamento verificadas nos modelos estimados, através da análise dos coeficientes das variáveis, foram confirmadas pelo cálculo dos valores médios dos grupos adimplentes e inadimplentes, não havendo nenhuma incoerência. Além disso, o teste t de diferença de médias evidenciou que as médias das variáveis explicativas nos dois grupos (adimplentes e inadimplentes) são estaticamente diferentes, a um nível de confiança de 95%. Considerou-se, portanto, que essa análise de perfil dos grupos, de acordo com as variáveis explicativas, foi suficiente para testar a validade dos modelos construídos.



#### **4.7 Possíveis contribuições dos modelos Credit Scoring para o Cred Cidadania**

Tomando por base os resultados já obtidos e demonstrados neste trabalho, informações e opiniões levantadas junto a profissionais da área creditícia no Cred Cidadania e a literatura sobre o microcrédito, foi realizada uma breve análise das possíveis contribuições dos modelos de Credit Scoring à instituição.

Foram identificadas, através da análise das informações supramencionadas, duas linhas principais de contribuição dos modelos para o Cred Cidadania, quais sejam, redução de custos operacionais e redução da inadimplência, que constituem, atualmente, dois grandes problemas para a instituição, afetando diretamente a sustentabilidade financeira da mesma.

Os altos custos operacionais da instituição decorrem de dois fatores principais: a personalização do processo de concessão do crédito e a reduzida escala operacional. A personalização do processo de concessão de crédito demanda, dos agentes e do comitê de crédito, tempo expressivo para acompanhar e avaliar os empréstimos. Em consequência disso, o número de clientes monitorados por cada agente é relativamente pequeno. Desta forma, os altos custos administrativos incorridos para manutenção da estrutura operacional da instituição não são satisfatoriamente diluídos, devido ao reduzido número de clientes, dificultando, assim, a obtenção de vantagens de escala na concessão de crédito.

Nesse contexto, a utilização de modelos Credit Scoring poderia fornecer algumas contribuições para a amenização do problema exposto. Os modelos agregam uma ampla gama de informações relativas à forma como as características do solicitante do crédito e do seu negócio afetam na inadimplência. Uma vez que a função principal do agente de crédito é justamente obter, diretamente nos empreendimentos dos clientes, dados que constituam insumos e forneçam respaldo à avaliação do risco, as informações provenientes dos modelos poderiam ser muito úteis nesse processo, visto que possibilitariam ao agente de crédito o conhecimento de algumas características dos clientes e dos seus empreendimentos, sem a necessidade de investigação em campo (diretamente com os clientes). As informações geradas pelos modelos contribuiriam, assim, para uma redução de tempo que os agentes de crédito dispõem na busca e análise de informações sobre os clientes, diminuindo, principalmente, o número de visitas realizadas com o objetivo de avaliar o potencial econômico dos empreendimentos e o emprego dos recursos emprestados. Com esse ganho de tempo, provavelmente, haveria uma ampliação da relação número de clientes para cada agente de crédito.

Além da economicidade de tempo, com a informação prévia fornecida pelos modelos sobre clientes mais propensos a se tornarem inadimplentes, os agentes de crédito poderiam

## CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE UMA INSTITUIÇÃO DE MICROCRÉDITO

direcionar maior atenção àqueles clientes que possuem maior tendência a serem maus pagadores, o que proporcionaria maior eficácia no desempenho de suas funções de acompanhamento prático dos créditos. A conjugação de maior eficiência dos agentes com redução de tempo dispensado pelos mesmos no acompanhamento e avaliação dos créditos resultaria, provavelmente, em um aumento de escala e redução de custos operacionais para a instituição.

Os modelos de Credit Scoring poderiam contribuir, também, para a redução das altas taxas de inadimplência da instituição. As informações disponibilizadas pelos modelos complementaríamos as análises dos agentes, provendo-os com insumos para uma atuação mais eficiente e eficaz na identificação do risco de inadimplência, conforme já mencionado. Estas informações também poderiam assessorar o comitê de crédito, capacitando-o a tomar decisões de concessão de crédito com um maior embasamento, de acordo com o nível de risco do cliente. Uma vez que as duas instâncias responsáveis pela análise e concessão do crédito na instituição estariam mais bem preparadas, em termos de informações, para realizar esta função, o processo de avaliação risco se tornaria, mais preciso, criterioso, com menor número de equívocos e um menor grau de subjetividade. Isso, sem dúvida, colaboraria para a redução da inadimplência no Cred Cidadania.

Além disso, com uso dos modelos seria possível a identificação prévia de probabilidade de ocorrência de inadimplência em cadeia dentro dos grupos solidários da instituição, questão que contribui consideravelmente para a alta taxa inadimplência.

Os problemas acima mencionados, quais sejam, os altos custos operacionais, as reduzidas escalas e a alta inadimplência da instituição, influenciam, maciçamente, na sustentabilidade financeira do Cred Cidadania. A fragilidade na auto-sustentação financeira constitui um grande entrave na maioria das instituições de microcrédito, não sendo diferente na instituição objeto deste trabalho. A não dependência de fontes externas de recursos para disponibilização em empréstimos e para cobertura de seus custos administrativos é um grande desafio que se coloca para a instituição, no curto e no longo prazo.

Assim, ao contribuir para redução de custos e inadimplência, os modelos de Credit Scoring estariam fornecendo expressivo auxílio para a sustentabilidade financeira da instituição no longo prazo.

## 5 Conclusões

Os modelos de risco de crédito desenvolvidos neste trabalho, tanto os de aprovação de crédito quanto aqueles behavioural scoring, apresentaram resultados próximos em ambas as técnicas estatísticas empregadas (análise discriminante e regressão logística), sendo similares quanto ao formato e quanto ao poder de previsão do risco de inadimplência ou poder de classificação dos clientes. Em média, todos os modelos classificaram 80% dos clientes da amostra corretamente, o que é considerado um resultado de classificação muito bom para modelos de credit scoring. No entanto, os modelos elaborados com regressão logística foram consideravelmente melhores na classificação correta dos clientes inadimplentes do que dos adimplentes. Uma vez que a previsão da inadimplência é considerada a principal finalidade dos modelos de risco de crédito, os modelos de regressão logística foram, portanto, mais eficazes no alcance de seus objetivos.

Nesse sentido, de uma forma geral, os resultados satisfatórios de previsão de inadimplência dos modelos obtidos neste trabalho indicam que, embora o microcrédito seja caracterizado com uma modalidade de crédito diferenciada, é possível a utilização de modelos de Credit Scoring na instituição de microcrédito estudada, como instrumentos de apoio ao processo de avaliação do risco de crédito. Devido às suas características, os modelos quantitativos de risco de crédito Credit Scoring podem fornecer contribuições relevantes para a amenização de alguns problemas da instituição, quais sejam, altos custos operacionais e altas taxas de inadimplência, problemas estes que exercem impacto direto na sustentabilidade financeira da mesma.

Ao contribuírem para a sustentabilidade financeira da instituição de microcrédito, os modelos de gestão de risco de inadimplência construídos neste trabalho podem, em consequência, fornecer expressivo subsídio ao alcance do objetivo primordial da instituição de microcrédito, qual seja, contribuir para o desenvolvimento econômico dos microempreendedores, através da concessão de crédito e geração de renda.

Considerando a possibilidade de realização e implementação de outros trabalhos dessa natureza, os resultados oriundos deste trabalho poderiam ser relevantes também para outras instituições de microcrédito, contribuindo para a evolução do conhecimento nesta área e fornecendo informações que dêem respaldo ao aperfeiçoamento das técnicas de gestão nestas instituições.

CONSTRUÇÃO DE MODELOS CREDIT SCORING COM ANÁLISE DISCRIMINANTE  
E REGRESSÃO LOGÍSTICA PARA A GESTÃO DO RISCO DE INADIMPLÊNCIA DE  
UMA INSTITUIÇÃO DE MICROCRÉDITO

REFERÊNCIAS

BARONE, F. M; LIMA, P. F; DANTAS, V; REZENDE, V. **Introdução ao Microcrédito**. Brasília: Conselho da Comunidade Solidária, 2002. Disponível em <<http://www.bndes.gov.br>>. Acesso em 01 dez. 2004.

BRUETT, T. **Manual de Técnicas de Gestão Microfinanceira: Programa de Desenvolvimento Institucional**. Rio de Janeiro: BNDES, 2002. <<http://www.bndes.gov.br>>. Acesso em 01 dez. 2004

CAOINETTE, J. ALTMAN, E; NARAYANAM, P. **Gestão do Risco de Crédito: o Próximo Grande Desafio Financeiro**. Rio de Janeiro: Qualitymark, 1999.

GUJARATI, D. N. **Econometria Básica**. 3. ed. São Paulo: Makron Books, 2000.

HAIR JÚNIOR, J. F; TATHAM, R. L; ANDERSON, R. E; BLACK, W. **Multivariate Data Analysis**. 5. ed. New Jersey: Prentice Hall, 1998.

KWITKO, E. E. B. (Coord.). **Manual para Formação de Agentes de Crédito: Programa de Crédito Produtivo Popular**. Rio de Janeiro: BNDES, 1999. Disponível em <<http://www.bndes.gov.br>>- Acesso em 23/12/2004.

LEWIS, E. **An Introduction to Credit scoring**. Fair Isaac: San Rafael, California. 1992.

SAUNDERS, Anthony. **Medindo o Risco de Crédito – Novas Abordagens para o Value at Risk e Outros Paradigmas**. Rio de Janeiro: Qualitymark, 2000.

SCHREINER, M. (1999) **A scoring Model of the Risk of Costly Arrears at a Microfinance Lender in Bolivia**. St. Louis, October, 1999. Disponível em <<http://www.microfinance.com>> Acesso em 15 fev. 2005.

THOMAS, L. C. A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Consumers, **International Journal of Forecasting**, Edinburgh -U.K, v. 16, p.149-172, 2000.